AD_____

AWARD NUMBER DAMD17-94-J-4443

TITLE: A Capillary Instrument for High Speed DNA-Based Tumor Diagnosis

PRINCIPAL INVESTIGATOR: Harold P. Swerdlow, Ph.D.

CONTRACTING ORGANIZATION: University of Utah
Salt Lake City, Utah 84102

REPORT DATE: September 1998

TYPE OF REPORT: Annual

PREPARED FOR: Commanding General
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1998	Annual (22 Aug 97 - 21 Aug 98)

4. TITLE AND SUBTITLE
A Capillary Instrument for High Speed DNA-Based Tumor Diagnosis

5. FUNDING NUMBERS
DAMD17-94-J-4443

6. AUTHOR(S)
Harold P. Swerdlow, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Utah
Salt Lake City, Utah 84102

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In order to test individuals for familial breast-cancer susceptibility, a fluidic 96-channel capillary DNA sequencing and analysis instrument is under development. The major subsystems of this instrument have been designed and tested. A novel manifold system will allow automated unattended loading to be performed. Ostensibly minor technical issues stand in the way of application of the prototype system to real samples on a large scale. User-friendly software is close to fully functional. The instrument will have major benefits over other commercial instruments that have been announced. Foremost is the ability to perform preparation, processing, loading, separation, detection, and analysis of DNA sequencing reactions in a coupled in-line fashion in a single instrument.

Simultaneously, techniques for the practice of capillary DNA sequencing are being refined. New matrices for separating DNA have been developed and optimized. Sample purity can be assessed using a novel miniature conductivity probe. A new technique for loading samples without purification has also been developed.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer diagnosis, polymerase chain reaction (PCR), capillary electrophoresis, automation, fluorescence, genotyping	13
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

HS Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

HS In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

HS In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Harold Sweddon 9/17/98

PI - Signature Date

Table of Contents

Introduction

Our previous annual report described the development of a 96-channel DNA sequencer using capillary electrophoresis (CE). This instrument will be validated for both breast-cancer diagnostics and for complete sequencing of the breast cancer-susceptibility loci. In the past year, since that report, two major commercial announcements have occurred in the field of capillary DNA sequencing. Molecular Dynamics' 96-channel capillary electrophoresis instrument (MegaBACE) has been officially introduced. Subsequently, a stunning proposal to commercially shotgun sequence the entire human genome using a new 96-channel capillary instrument (model 3700) has been made by Perkin-Elmer/Applied Biosystems and The Institute for Genome Research (TIGR). To the casual observer, it might seem prudent to shut down any instrument development efforts in this arena.

However, there are serious questions that arise concerning these other instruments. Some of them were touched upon in the previous report re the Molecular Dynamics instrument, e.g., unattended operation, polymer viscosity/read-length, problems with the scanning approach, flexibility of the optical design for new dye sets. Reports from the field have been mixed with respect to reliability and read-length for the MegaBACE. Some of these problems will also impact the as yet unseen 3700 instrument.

For these reasons we have not retreated, choosing instead to refine our strategy. We continue to develop our 96-capillary detector. However, we have increased our efforts in other areas. The instruments that have been announced make no attempt to provide fluidic front-end loading strategies, a concept we still firmly believe in. Fluidic loading is not specific to our instrument - it will have applicability to any detection/instrument design. We are still hard at work on this aspect of capillary DNA sequencing. Additionally, we have begun to push harder on improving the reliability and simplicity of CE technology for DNA sequencing. We are trying to learn more about running single capillaries, while simultaneously preparing for the 96-capillary leap. These efforts are concentrated in polymer-buffer optimization and sample preparation. Sample preparation is much more difficult to perform and to quality control for capillaries than it is for slab gels. In my opinion, sample preparation issues, not instrument price or features, will determine whether CE instruments attain a significant market share for large, high-throughput DNA sequencing projects.

Some of this work has been left unfinished, due in large part to the fact that my laboratory, but none of the personnel, has relocated to the newly formed Center for Genomics Research at the Karolinska Institute in Stockholm, Sweden. This move, although severely disruptive at present, is expected to greatly increase the quality and quantity of output from my laboratory in the future.

Body

We will organize this report into four sections, generally in line with the format from last year's report.

1) Flushing and refilling the capillaries

Capillaries will be refilled from a positive-buffer-chamber pump that was designed and built in-house. The basic strategy is to attach all 96 capillaries through a manifold into the exit port of a high-pressure (1000 psi) piston pump. The pump is modular with respect to the number of capillaries, and allows the use of multiple buffers (e.g., running and refill buffers). The pump is electrically isolated, so that after refilling, the capillaries can be connected to the buffer reservoir directly. The high-voltage power supply can deliver 20,000 volts through this pump without shorting to ground. Pressure and electrical testing and flow studies have now been completed, and the device is ready for installation into the automated sequencing system.

The successful implementation of a multiple-capillary system relies critically upon a good separation matrix and stable capillaries. This is a difficult problem, but one we have relentlessly pursued. Polymer-filled capillaries containing either poly(dimethylacrylamide) pDMA, or a copolymer of allyl-b D glucopyranoside and acrylamide pAGAA, have been tested in our laboratory. Both polymers have very low viscosity (100-700 cp). This attribute will allow very short refill times (1-2 min). Low viscosity almost certainly leads to reliable and reproducible refilling, preventing clogging and providing a uniform separation matrix, essential for good resolution. Both polymers are also self-coating; i.e., they require no capillary wall pre-treatment. Self-coating polymers allow the system to be completely replenished after electrophoresis; sequential multiple injections occur with no complicated washing steps between runs. The replenishment step severely reduces problems with capillary performance degradation that occur with covalently-attached wall coatings. Self-coating polymer buffer systems are also less susceptible to sample-induced clogging and fouling of the capillary surface. Therefore, sample preparation may be simpler in these systems.

The pDMA we obtained from PE/Applied Biosystems has been optimized here to work reliably and stably, providing 650 base reads in 2 hours with at least 100 runs per capillary (Xiong, Y., Wang, H., Nay, L., and Swerdlow, H. "Improvements in the Use of an Entangled-Polymer Buffer, pDMA, for Capillary Electrophoresis of DNA Sequencing Reactions" - in preparation). pAGAA-containing buffers (developed under a collaboration with Dr. Marcella Chiari at the Istituto di Chimica degli Ormoni, in Milan, Italy) performed almost identically, giving typical read-lengths of 630 bases in 2 hours. This work, will be submitted very soon under the title "A Novel Capillary Electrophoresis Sieving Polymer Matrix for DNA Sequencing: Performance and Optimization," by Haibo Wang, Ceriotti Laura, Marcella Chiari, and Harold Swerdlow.

2) Loading DNA onto the capillaries

The effort to develop a fluidic-loading device has proceeded well. This new manifold loader represents a major redesign of the device used in our original fluidic analysis system, described in "Fully Automated DNA Reaction and Analysis in a Fluidic Capillary Instrument" (Analytical Chemistry, vol. 69, 1997 pp. 848-855). Both loading strategies are based on a junction between a separation capillary and a fluidic path. In the newest design, we employ a sipping tube as the fluidic path. The sipper transports the sample from, e.g., a microtiter dish, and positions the sample droplet where it can be loaded electrokinetically onto the capillary. This "immersed-droplet" method can be easily adapted to multiple-capillary loadings. We create all the (96) junctions within a single chamber or manifold, allowing all the valves and pressure sources to be located externally to the manifold loader. The device has been thoroughly tested with single-capillary separations of size marker samples; resolution and signal are as good as conventional loading techniques. An 8-channel prototype and associated protocol are currently being optimized with DNA-sequencing reactions. The system has been made more robust and easier to machine recently, by incorporation of a sleeve designed to hold the capillary and sipper tube in precise registration. The fluidics have been redesigned to be compatible with computer control of all valves and pumps. The current prototype can handle up to 16 capillaries, and a 96-channel prototype is only a few keystrokes away. The entire device has been designed and built on a computer-aided manufacturing workstation. We are confident that such fluidic-loading strategies will be the cornerstone of both our group's and other's efforts to automate the process of electrophoresis for high-throughput environments.

We are also pursuing an on-column sample concentration method for capillary-based DNA sequencing. This base-stacking method allows direct injection of unpurified products of dye-primer sequencing reactions onto capillaries without any pre-treatment. Such a system has the potential to eliminate tedious and labor-intensive sample purification steps required for capillary sequencing beyond those already needed for slab-gel based methods. Concentration of samples is also achieved, resulting in higher signals for both normal samples and large-volume dilute samples. On-column concentration of DNA fragments is achieved simply by electrokinetic injection of hydroxide ions using a novel Tris/HCl separation buffer. A neutralization reaction between these OH- ions and the cationic buffer component Tris+ results in a zone of lower conductivity, within which field focusing occurs. DNA fragments are concentrated at the front of this low conductivity zone. Figure 1 shows a comparison of conventional loading of purified samples, conventional loading of unpurified samples, and base stacking of unpurified samples. Both the resolution and signal strength seen for base-stacked crude samples are equivalent to that obtained from conventional injections of highly purified samples.

With sample injection times as long as 360 s at 50 V/cm, resolution could still be restored by the stacking process, with only minor loss of resolution. Using a 36 cm effective length uncoated capillary, with poly(dimethylacrylamide) as the separation matrix, and an electric field of 160V/cm, a resolution of at least 0.5 could be obtained for fragments up to 650 nucleotides long. No significant degradation of the capillary performance was observed over at least 20 sequencing runs using this new sample injection method. Unlike ethanol precipitation, the method is quick, simple, and can be easily performed on conventional CE instruments without

hardware modification. The technique has the potential to aid in the development of fully-automated multi-capillary sequencing systems.

Dye-primer sequencing samples could be injected directly onto polymer-filled capillaries using base-stacking. However, at present, the base-stacking method does not perform well with crude samples generated using dye-terminator sequencing chemistry. The dye-terminator reactions suffer from large interfering dye peaks that are normally eliminated by ethanol precipitation. We are currently trying to overcome this limitation.

The base-stacking technique can be useful even for purified samples in cases where dilution is significant and/or where signal strength is particularly poor. Certain DNA analysis methods can benefit from injection of large sample volumes without loss of resolution due to overloading. Future experiments will test whether this technique can be easily adapted for concentrating large volume samples onto slab gels, without loss of resolution.

The details of this technique have been accepted for publication in *Analytical Chemistry*, under the title "Base Stacking: pH-Mediated On-Column Sample Concentration for Capillary DNA Sequencing," by Yan Xiong, Sang-Ryoul Park, and Harold Swerdlow.

In another study (manuscript in preparation), loading efficiency of DNA sequencing reaction products for capillary electrophoresis was improved by using low-ionic-strength solvents to resuspend the samples. This technique will be important for reliably obtaining good signal strengths in capillary DNA sequencing runs. Transference number (fraction of charge carried by the sample constituents) of fluorescent primers and sequencing reaction products was determined. This number is conceptually very similar to the conductivity due to the DNA (being analyzed) only. This parameter was quantitated by measuring the conductivity of real DNA sequencing samples (volumes of only a few microliters), using a novel miniature conductivity probe we developed. A simple quantitative model for sample loading was also derived. According to this model, the amount of DNA loaded by electrokinetic injection depends directly on the ion content of the DNA sample, and indirectly on the ion content of the sample solvent. The predictions of this model have been experimentally verified.

Treating sample solvents with mixed-bed resin results in a significant decrease in the conductivity of the solvent. The use of these low-conductivity solvents to dissolve cleaned-up DNA sequencing samples results in significantly increased efficiency of injection of the DNA onto polymer-filled capillaries. Improved loading will be important for large-scale sequencing projects and large-scale diagnostic efforts like BRCA 1 and 2 screening, due to savings in both reagents and template.

The use of this novel conductivity probe also helped address another issue. In capillary electrophoresis of DNA, the fraction of a sample that is loaded electrokinetically is usually difficult to calculate or measure. It has long been thought that DNA sequencing reactions can be quantitatively loaded onto capillaries when low-ionic-strength sample buffers are employed. Quantitative loading would imply that significant depletion of samples will occur with repeated injections.

By using our miniature probe (figure 2), the conductivity of a DNA sample (squares) and that of a control containing only sample matrix (diamonds) were measured after each injection

onto a polymer-filled capillary. With repeated injection, the total conductivity of both samples increases by the same amount (figure 2). By subtracting the two curves, the conductivity due to the DNA alone can be obtained (triangles), and it is constant. It is clear that the lower signal seen after multiple injections of the same sample was due to an increase in salt in the sample, and not due to a depletion of DNA. This result implies that even under conditions where the salt concentration in a sample matrix is very low, only a very small portion of a typical DNA sample is loaded during an electrokinetic injection onto a capillary. Contamination of the sample with salt occurs despite dipping the capillary repeatedly. The samples became contaminated either from running-buffer ions left on the capillary outer wall or directly from polymer-matrix ions leaching from inside the capillary. Since extensive capillary washes were employed by the automated CE instrument we used, the latter explanation is preferred. A manuscript describing this work is currently being prepared for publication.

3) Detection and operation of the capillary instrument

We have constructed an imaging detection system that has no moving parts, employs full-spectral detection of multiple fluorescent dyes, has very high sensitivity due to excellent light collection for all capillaries and is easy to align and use. This system is based upon fiber-optic excitation and charge-coupled-device (CCD) camera detection of fluorescent light. We have now employed new optimized components (larger lenses, new grating, new CCD camera) and thoroughly tested them. We have also had to make a choice between a commercially-available fiber-optic splitter and a home-made version. This choice was made simpler by the recent availability of miniature translation stages from Newport Optical (Irvine, CA). These stages will allow us to micro-position the different elements of the commercial fiber-optic splitter device on a single platform. This new design will allow optimal alignment with no noise due to fluorescence in glues and fittings. The detector has been tested in an 8-capillary mode, although 96 capillary detection is already possible. The key problem we have faced for some time is a loss of resolution when this detector is employed with multiple capillaries. We believe this problem to be electrical in nature, due to the very thin walls of our capillaries and the fact that they are held in place on a conductive silicon substrate. However, numerous tests to find the source of this problem have been inconclusive to date, and the project has been hampered significantly by the relocation of my laboratory. Tracking down this problem will be a primary goal of the new laboratory team.

This loss of resolution has complicated the deployment of the multiple-capillary fluidics manifold since it is difficult to determine if the fluidics subsystems are performing as expected. However, once the first problem has been tracked down, we expect the development to proceed rather smoothly.

4) Data collection and analysis

As with most new technological advances, the success of capillary technology will depend to a large extent on the quality of the instrument software. To this end, we have worked hard to formulate reliable algorithms and user-friendly tools for our system. Pre-run images of the capillary array are transferred from the CCD camera to a high-end workstation, where they are dynamically displayed; brightness, contrast, and size can be adjusted at will. A set of tools

has been created for manipulating these images, including a sophisticated "striping" tool that will allow users to quickly find the capillaries and the spectra on the pre-run images. This tool can create 1 or many (96) boxes at a time. Boxes can be repositioned and resized, together or individually, and can be saved for subsequent experiments. Coordinates for spectral alignment within each box are obtained at the same time from the Rayleigh-scattered laser-line position. The box coordinates are transferred to the CCD camera controller to provide summing instructions that tell the controller how to add individual elements (pixels) on the chip to obtain individual capillary spectral data. This process is referred to as binning. Binning has the advantage that noise is reduced relative to off-chip summing and it can be performed much faster (in a matter of microseconds). For each time point during the run, images are binned and the spectra obtained are aligned to each other using the spectral alignment coordinates saved previously. In this way, all 96 capillary spectra begin and end in the same place and each contains the same number (about 100) points. This is key for the successful application of a single 4x100 "inverse" matrix for all capillaries that converts the spectral data to four color "trace" data. The massive reduction in data that is obtained in real time by binning (about 50 fold) and matrix inversion (about 25 fold), is key to the routine use of this instrument. File sizes for the entire run are brought within reasonable limits. The obtained trace data is conceptually identical to the input specifications for all standard base-calling software packages. However, in some cases, signal processing may be necessary to reduce noise specifically for this application. Furthermore, mobility shift correction may be required, if not already accounted for in the base-caller. Mobility shift correction is necessary to reverse the electrophoretic effect of the different fluorescent dyes attached to the different fragments (A, T, G, or C). The design and testing of the base-calling software is being done elsewhere, and is beyond the scope of the present project.

Conclusions

We have made good progress towards development of our 96-capillary sequencer. We are now able to reliably detect signals from 8 capillaries with good signal-to-noise characteristics. The optics for 96-capillary detection are in place. The two fluidic subsystems, one for refilling capillaries and the other for automated in-line sample loading, are ready to be incorporated into the 96-channel prototype, as soon as some technical difficulties are surmounted. Our instrument will have significant advantages over other instruments currently out or announced on the market. Most importantly, this instrument has the ability to prepare and process DNA in-line in a single instrument. This will afford the user far-reduced handling time and will free up technical staff for the more interesting aspects of DNA analysis. Furthermore, it is highly flexible, being easily adapted to either short fast runs for diagnostic sequencing (looking for known mutations in BRCA1 and 2, for example), or to detection of multiple dyes for high-throughput genotyping. Software being developed for this instrument is easy to use and could be useful for similar instruments being designed.

We have succeeded in co-developing a separation matrix that has low-viscosity, high resolving power, and works dynamically in uncoated capillaries. This copolymer of allyl-b D glucopyranoside and acrylamide is a viable alternative to the poly(dimethylacrylamide) formulation being marketed by Perkin-Elmer/Applied Biosystems. Having such high quality polymers in the public domain will be a great boon to independent capillary instrumentation developers and users.

Another generally useful contribution to the field of capillary DNA sequencing, is the fact that we have refined procedures for sample preparation and processing. Our base stacking procedure has great potential to eliminate tedious purification steps normally preformed after DNA sequencing reactions are complete. Base stacking will likely be very useful also for dilute samples, and those for which high signal strength is important. Sample purity issues have continually plagued capillary sequencing methods since their introduction at the beginning of this decade. Examination of sample purity by inexpensive direct means (using our novel micro-conductivity probe) thus has enormous benefits for rigorous quality control in DNA sequencing operations. In fact, such a procedural check is essential for routine application of CE methods in large-scale sequencing tasks. Similarly, addressing questions about sample depletion during repeated injections is not only of academic interest, as it impacts efficiency of sample loading for all analyses.

The immediate goal of this work should be attainable within the next year of funding. A fully-automated fluidic capillary electrophoresis instrument that can prepare, process, load, separate, detect, and analyze DNA sequencing reactions would be a boon for DNA-based breast cancer diagnostics and many other applications.

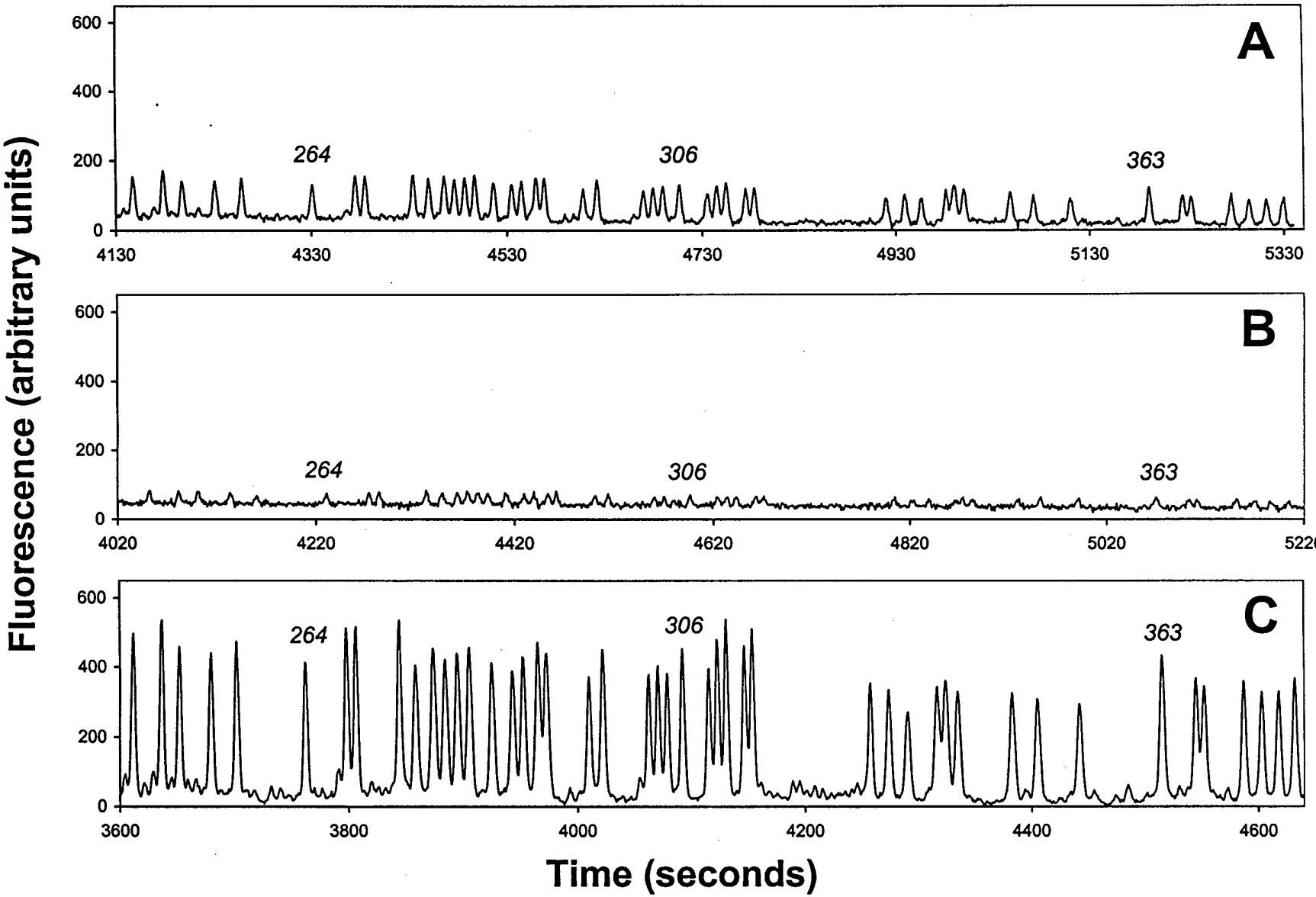

Figure 1. Comparison of conventional sample injection with base stacking. (A) A portion of the electropherogram of a single-terminator, dye-primer, DNA sequencing reaction - "purified sample" - injected for 30 sec at 50 V/cm using "TAPS/NaOH buffer". (B) Untreated "crude sample" injected as in A. (C) "Crude sample" injected for 130 sec at 50 V/cm using "Tris/HCl buffer"; base stacking was achieved by injection of NaOH (0.1N) for 60 sec at 160 V/cm. Samples were generated from M13mp18 DNA using FAM-labeled -21 Universal Primer, ddGTP, and ThermoSequenase. A 50 mm i.d., 375 mm o.d., 36 cm effective length, uncoated capillary was used. Electrophoresis was performed on an ABI 310 instrument operating at 160 V/cm and 42 °C. Peaks correspond to Gs in the known sequence; peak numbers refer to total nucleotide length of the analyzed fragments

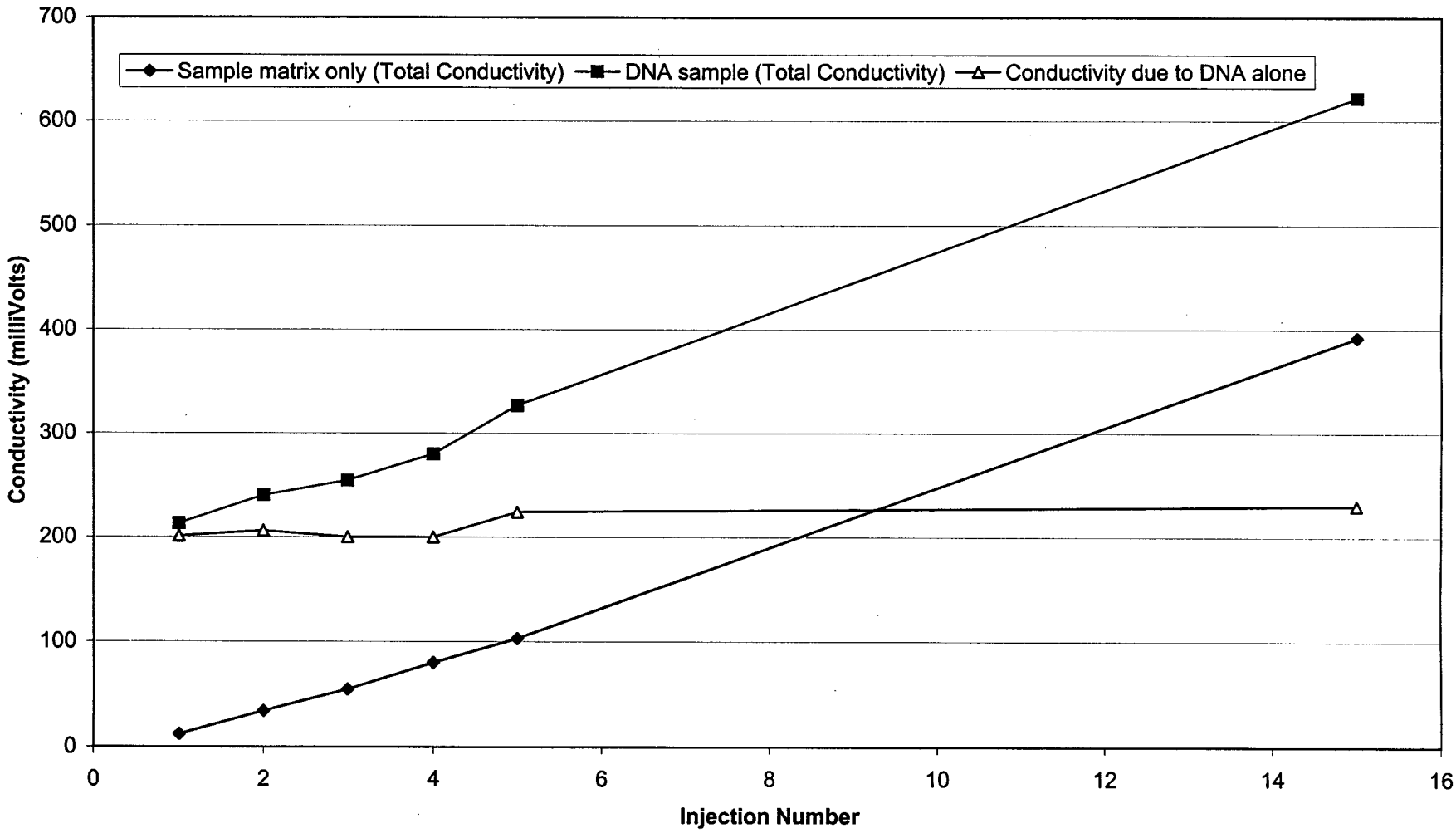

Figure 2. Conductivity measured directly for different samples prior to repeated injection onto a capillary. The curve for "conductivity due to DNA alone" is obtained by subtracting the two other curves,